MIS Quarterly

RESEARCH ARTICLE

ENGAGING VOLUNTARY CONTRIBUTIONS IN ONLINE COMMUNITIES: A HIDDEN MARKOV MODEL

Wei Chen

Eller College of Management, University of Arizona, Tucson, AZ 85721 U.S.A. {weichen@email.arizona.edu}

Xiahua Wei

School of Business, University of Washington, Bothell, WA 98011 U.S.A. {xhwei@uw.edu}

Kevin Xiaoguo Zhu

Rady School of Management, University of California, San Diego, CA 92093-0553 U.S.A. {kxzhu@ucsd.edu}

Appendix A

The MCMC Estimation of the HMM

We estimate the parameters vector $\{\theta, \tilde{S}\}$ with Gibbs sampling (Albert and Chib 1993). Suppose we have motivation state $s_{it} \in \{1, 2, \ldots, J\}$ in our model. We generate the joint posterior distribution by sampling from each conditional distribution of the following parameter blocks:

$$\begin{aligned}
&\boldsymbol{\theta} = (\theta_1, \theta_2', \theta_3', \theta_4', \theta_5')' \\
&\theta_1 = \sigma^2 \\
&\theta_2 = (\beta_1, \beta_2', \ldots, \beta_J')' \\
&\theta_3 = (\xi_1, \xi_2', \ldots, \xi_J')' \\
&\theta_4 = (\mu_2, \mu_3, \ldots, \mu_{J\text{-}1})' \\
&\theta_5 = (L_{i1}, L_{i2}, \ldots, L_{iT})',\ i = 1, \ldots, n \\
&\theta_6 = (s_{i1}, s_{i2}, \ldots, s_{iT})',\ i = 1, \ldots, n
\end{aligned}$$

For the simplicity of presentation, we denote $\theta_{-i} = (\theta_j')'$, $\forall j \neq i$ below.

(1) Sample $\theta_1 = \sigma^{-2}$ **from** $P(\theta_1|\theta_{-1}, Y, X, W)$.

Prior: $\sigma^{-2} \sim \Gamma(\alpha, \delta)$. Conditional on θ_{-1}, Y, X, and W, it is equivalent to observing $\{\varepsilon_{it}\}$ where $\varepsilon_{it} = Y_{it} - X_{it}\beta_{s_{it}}$.

Posterior: $(\sigma^{-2}|\theta_{-1}, Y, X, W) \sim \Gamma(\alpha + \frac{1}{2}\, nT, \delta + \frac{1}{2}\, SSR)$, where $SSR = \sum_{i=1}^{n}\sum_{t=1}^{T}\varepsilon_{it}^2$.

(2) Sample $\theta_2 = (\beta_1, \beta_2', \ldots, \beta_j')'$ **from** $P(\theta_2|\theta_{-2}, Y, X, W)$.

Prior: $(\beta_1|\sigma^{-2}) \sim N(m_j, M_j)$, $j = 1, \ldots, J$ (independent of each other)

Posterior: Conditional on $\{s_{it}\}$, only those observations for which $s_{it} = j$ are relevant to posterior distribution of β_j: $(\beta_j|\theta_{-2}, Y, X, W) \sim N(m_j^*, M_j^*)$, where

$$* M_j^* = \left(M_j^{-1} + \sigma^{-2} \sum_{i=1}^{n} \sum_{t=1}^{T} X_{it} X_{it}' \, 1_{\{s_{it}=j\}} \right)^{-1}$$

and

$$m_j^* = M_j^* \left(M_j^{-1} m_j + \sigma^{-2} \sum_{i=1}^{n} \sum_{t=1}^{T} X_{it} Y_{it} \, 1_{\{s_{it}=j\}} \right)$$

(3) Sample $\theta_3 = (\xi_1', \xi_2', \ldots, \xi_J')'$ **from** $P(\theta_3 | \theta_{-3}, Y, X, W)$.
Prior: $\xi_j \sim N(mw_j, Mw_j)$, $j = 1, \ldots, J$.
Posterior: $(\beta_j | \theta_{-3}, Y, X, W) \sim N(mw_j^*, Mw_j^*)$, where

$$Mw_j^* = \left(Mw_j^{-1} + \sum_{i=1}^{n} \sum_{t=1}^{T} W_{i,t-1} W_{i,t-1}' \, 1_{\{s_{i,t-1}=j\}} \right)^{-1}$$

and

$$mw_j^* = Mw_j^* \left(Mw_j^{-1} mw_j + \sum_{i=1}^{n} \sum_{t=1}^{T} W_{i,t-1} L_{it} \, 1_{\{s_{i,t-1}=j\}} \right)$$

Note that since σ_u^2 is not identifiable, we normalize it to 1 in the estimation.

(4) Sample $\theta_4 = (\mu_2, \mu_3, \ldots, \mu_{J-1})'$ **from** $P(\theta_4 | \theta_{-4}, Y, X, W)$.
Albert and Chib provide the posterior for μ_j given the other threshold parameters μ_k, $k \neq j$. For each μ_j, let $Lower = \max\{\max\{L_{it}: s_{it} = j\}, \mu_{j-1}\}$ and $Upper = \min\{\min\{L_{it}: s_{it} = j+1\}, \mu_{j+1}\}$. Then we can sample μ_j from the uniform distribution $U[Lower, Upper]$.

(5) Sample $\theta_5 = (L_{i1}, L_{i2}, \ldots, L_{iT})', i = 1, \ldots, n$ **from** $P(\theta_5 | \theta_{-5}, Y, X, W)$.
L_{it} determines s_{it} according to the following formula: $s_{it} = j \text{ if } \mu_{j-1} < L_{it} < \mu_j$, where $\mu_0 = -\infty$, $\mu_1 = 0$, $\mu_J = \infty$, and $\mu_2, \ldots, \mu_{J-1}$ are given in step (4). Conditional on θ_{-5}, we can generate L_{it} from a truncated normal distribution $TN_{(\mu_{j-1}, \mu_j)}(W_{i,t-1}\xi_{s_{i,t-1}}, 1)$, which is a normal distribution with mean $W_{i,t-1}\xi_{s_{i,t-1}}$ and variance 1, and truncated left at μ_{j-1} and right at μ_j. Repeating this for $t = 1, \ldots, T$ and $i = 1, \ldots, n$ gives a draw from $P(\theta_5 | \theta_{-5}, Y, X, W)$.

(6) Sample $\theta_6 = (s_{i1}, s_{i2}, \ldots, s_{iT})', i = 1, \ldots, n$ **from** $P(\theta_6 | \theta_{-6}, Y, X, W)$.
We generate the states using the single-move Gibbs-sampling algorithm in Kim and Nelson (1999), which is also the well-known Forward-Backward algorithm. Denoting Ψ_{it} as information for user i up to time t, and Ψ_{iT} as information from the whole sample, we follow the forward-backward algorithm as below to obtain $P(s_{it} | S_{i,-t}, \Psi_{iT})$:

(a) **Forward**: Calculate $P(s_{it} | \Psi_{it})$.
Step 1: Given $P(s_{i,t-1} = k | \Psi_{i,t-1})$, $k = 1, \ldots, J$ at the beginning of period t, calculate $P(s_{it} = j, s_{i,t-1} = k | \Psi_{i,t-1}) = P(s_{it} = j | s_{i,t-1} = k, \Psi_{i,t-1}) P(s_{i,t-1} = j | \Psi_{i,t-1})$, where

$$P(s_{it} = j | s_{i,t-1} = k, \Psi_{i,t-1}) = \begin{cases} \Phi(\mu_1 - W_{i,t-1}\xi_k), & if\ j = 1 \\ \Phi(\mu_j - W_{i,t-1}\xi_k) - \Phi(\mu_{j-1} - W_{i,t-1}\xi_k), & if\ j = 2, \ldots, J-1 \\ 1 - \Phi(\mu_{j-1} - W_{i,t-1}\xi_j), & if\ j = J \end{cases}$$

For the first period, we use the initial probability $P(s_{i1} = j) = p_j$ for $j = 1, \ldots, J$, which are sampled from a Dirichlet distribution.
Step 2: Once X_{it} and Y_{it} are observed in period t, we update the probability term by calculating $P(s_{it} = j | \Psi_{it}) = \sum_{k=1}^{J} P(s_{it} = j, s_{i,t-1} = k | \Psi_{it})$, where

$$
\begin{aligned}
P\big(s_{it} = j, s_{i,t-1} = k \mid \Psi_{it}\big) \\
&= P\big(s_{it} = j, s_{i,t-1} = k \mid \Psi_{i,t-1}, X_{it}, Y_{it}\big) \\
&= \frac{f\big(Y_{it} \big| s_{it} = j, s_{i,t-1} = k, \Psi_{i,t-1}, X_{it}\big) P\big(s_{it} = j, s_{i,t-1} = k \big| \Psi_{i,t-1}\big)}{f\big(Y_{it} \big| \Psi_{i,t-1}, X_{it}\big)} \\
&\propto f\big(Y_{it} | s_{it} = j, X_{it}\big) P\big(s_{it} = j, s_{i,t-1} = k \big| \Psi_{i,t-1}\big)
\end{aligned}
$$

(b) **Backward**: In the backward process, we generate s_{it} conditioning on Ψ_{it} and $s_{i,t+1}$ $(t = T-1, T-2, \dots, 1)$ using $g(s_{it}|\Psi_{it}, s_{i,t+1}) \propto g\big(s_{i,t+1}\big|s_{it}, \Psi_{it}\big) g(s_{it}|\Psi_{it})$. We then can calculate

$$
P\big(s_{it} = j \big| s_{i,t+1}, \Psi_{it}\big) = \frac{g\big(s_{i,t+1}\big|s_{it} = j, \Psi_{it}\big) g(s_{it} = j|\Psi_{it})}{\sum_{k=1}^{J} g\big(s_{i,t+1}\big|s_{it} = k, \Psi_{it}\big) g(s_{it} = k|\Psi_{it})}
$$

Then we can use a random number drawn from a uniform distribution to generate s_{it} according to $P(s_{it}|S_{i,-t}, \Psi_{iT})$.

Appendix B

Log-Likelihood and Model Selection Criteria

As detailed in Appendix A, we estimate the parameters in our HMM with Bayesian estimation, which does not require us to calculate the likelihood. However, to select the number of states, the selection criteria would rely on the likelihood. Therefore, we describe the calculation of the likelihood of an observed sequence of contributions and the selection criteria below.

Log-Likelihood Calculation

Because we adopt a hidden Markov model, the contribution probabilities for each individual over time are correlated through the hidden states. The joint likelihood of each individual's contribution sequence has to consider the possible paths of the underlying states (Netzer et al. 2008). Suppose that there are J possible states. Then according to MacDonald and Zucchini (1997), we can write the joint probability using a matrix product as

$$
P_i(Y_{i1} = y_{i1}, \dots, Y_{iT} = y_{iT}) = P_0 \Omega_i(1) Q_i(1,2) \Omega_i(2) \cdots Q_i(T-1,T) \Omega_i(T) \mathbf{1}'
$$

where P_0 is the initial probability, $\Omega_i(t)$ is a $J \times J$ diagonal matrix with the elements of emission probability $\omega_{it|j} = f(Y_{it}|X_{it}, s_{it} = j; \boldsymbol{\beta}, \sigma^2)$ on the diagonal, $Q_i(t-1,t)$ is the $J \times J$ transition matrix for individual i at time t with the elements of $q_i(k,j) = f\big(s_{it} = j \big| W_{i,t-1}, s_{i,t-1} = k; \boldsymbol{\xi}\big)$ on the k^{th} row and j^{th} column, and $\mathbf{1}'$ is a $J \times 1$ vector of ones. The element probabilities are obtained according to our model setup:

$$
\omega_{it|j} = f(Y_{it}|X_{it}, s_{it} = j; \boldsymbol{\beta}, \sigma^2) = \left\{1 - \Phi\left(\frac{X_{it}'\beta_j}{\sigma}\right)\right\}^{1\{Y_{it}=0\}} \left\{\frac{1}{\sigma}\phi\left(\frac{Y_{it} - X_{it}'\beta_j}{\sigma}\right)\right\}^{1\{Y_{it}>0\}}
$$

and

$$
q_i(k,j) = f\big(s_{it} = j \big| W_{i,t-1}, s_{i,t-1} = k; \boldsymbol{\xi}\big) = \Phi\big(\mu_{j+1} - W_{i,t-1}\xi_k\big) - \Phi\big(\mu_j - W_{i,t-1}\xi_k\big)
$$

Then we can write the log-likelihood as $\ln L = \sum_i \log(P_i)$.

Selection Criteria

We adopt three model selection criteria to determine the number of states in our HMM. First, we use the commonly used Akaike information criterion (AIC) and Bayesian information criterion (BIC) (Singh et al. 2011; Yan and Tan 2014):

$$AIC = -2 * \ln L + 2 * size$$

and

$$BIC = -2 * \ln L + size * \ln N$$

where *size* is the number of parameters in the model, and *N* is the number of users in the sample. Second, realizing that we are using a Bayesian estimation for our HMM, we also adopt Markov switching criterion (MSC), which was developed for HMM's state and variable selection (Smith et al. 2006). We follow the adaptation in the literature for its formulation (Netzer et al. 2008):

$$MSC = -2 * \ln L + \sum_{s=1}^{J} \frac{\hat{T}_s\left(\hat{T}_s + J * K\right)}{\hat{T}_s - J * K + 2}$$

where $\hat{T}_s = \sum_{t=1}^{T} \sum_{i=1}^{N_t} P(s_{it} = s)$, J is the number of states in the model, and K is the number of covariates in both the transition matrix and the state-dependent vector.

Appendix C

Testing the Estimation on Simulated Data

Because our model has a nonlinear feature by incorporating the *Tobit* and *probit* models, we could not use standard statistical software to estimate it. We have to write our own estimation algorithm instead. Hence we did, but we need to ensure that it is correct before applying the algorithm to the actual data. We run the algorithm on simulated data based on known parameters, and test whether it could recover the true parameters. Because there is some model uncertainty on the number of states in our HMM, we also simulate data with 2, 3, and 4 true states, and then estimate the model with 2, 3, and 4 states in HMM. Then we use the model selection criteria to determine whether our algorithm points out the true number of states. Here we use three true states as an example.

We first generate the true parameters $\boldsymbol{\theta}$, the community and individual characteristics variables $X = \{X_{it}\}_{t=1,\ldots,T;i=1,\ldots,N_t}$, and the community interaction variables $W = \{W_{it}\}_{t=1,\ldots,T;i=1,\ldots,N_t}$ with three motivation states ($J = 3$). Since we assume that a user has an initial probability $P_0 = \{p_1, p_2, p_3\}$, at $t = 1$ we draw the initial state s_{i1} of user i from a Dirichlet distribution using the initial probability P_0 for each user i that enters the community. Conditional on s_{i1}, we then draw the contribution $Y_{i1} = \max(0, Y_{i1}^*)$, where $Y_{i1}^* = X_{it}\beta_{s_{i1}} + \varepsilon_{i1}$ and ε_{i1} is generated from a normal distribution with mean 0 and variance σ^2. For any $t > 1$, we first draw $L_{it} = W_{i,t-1}\xi_{s_{i,t-1}} + u_{it}$, where u_{it} is drawn from $N(0,1)$. Then we generate the new state s_{it} according to L_{it}. Repeating the same process, we generate $Y = \{Y_{it}\}_{t=2,\ldots,T;i=1,\ldots,N_t}$ for all t.

With the simulation data {X, W, Y}, we estimate the model with our procedure and present the results in Table C1. Our simulation data contains 322 individuals and 20 periods of time. The true number of states is $J = 3$. The community and individual characteristics vector X contains four variables, and the community interaction vector W contains four variables. In Table C1, the "True Parameters" panel on the left displays the original parameters $\theta = \{\beta, \xi, \sigma^2\}$ that we employ to generate the simulation data. The "Estimation" column on the right displays the estimated parameters. Our estimation recovers the true parameters accurately.

We also present the model selection criteria in Table C2. Given the true state number is three, all our model selection criteria indicate that our HMM model with three states fit the data the best. This confirms the reliability of the estimation algorithm, and gives us confidence in its empirical application to the actual data.

Table C1. Estimation Results from Simulation Data (Number of States = 3)						
	True Parameters			Estimation		
Variables	State 1	State 2	State 3	State 1	State 2	State 3
$\boldsymbol{\beta}$				Mean (Standard Deviation)		
x_1	3	5	7	2.98 (0.07)	4.99 (0.07)	6.99 (0.03)
x_2	4	6	8	4.00 (0.02)	6.02 (0.01)	8.01 (0.01)
x_3	5	7	9	4.99 (0.02)	7.00 (0.02)	8.98 (0.01)
x_4	6	8	10	6.00 (0.02)	8.00 (0.01)	10.01 (0.01)
σ^2	1.5			1.53 (0.03)		
$\boldsymbol{\xi}$						
w_1	-1.5	-0.5	0.5	-1.63 (0.13)	-0.48 (0.08)	0.42 (0.06)
w_2	1.15	0.37	2.53	1.18 (0.11)	0.27 (0.08)	2.57 (0.14)
w_3	6.32	4.48	7.35	6.53 (0.29)	4.41 (0.22)	7.67 (0.44)
w_4	2.65	3.05	6.96	2.73 (0.15)	3.10 (0.12)	7.04 (0.17)
μ_j		2			1.95 (0.04)	
σ_u^2	1			1		
$P_0 = \{p_j\}$	0.45	0.40	0.15	0.44 (0.02)	0.41 (0.03)	0.15 (0.020)
T = 20	*N* = 322	Draws = 2,000				

Table C2. Selection of Number of States from Simulation Data (Number of States = 3)					
Number of States	**- 2*Log-likelihood**	**AIC**	**BIC**	**MSC**	**Number of Variables**
2	35404.87	35442.87	35514.59	41380.24	19
3	**22700.68**	**22758.68**	**22868.14**	**22700.68**	**29**
4	22734.37	22812.37	22959.57	30211.52	39

Appendix D

Robustness Checks

We conduct several sets of robustness checks. First, we estimate the model on another sample period (301-500 days). The results are in Table D1. Second, we examine whether the moderator role of a user or new questions by the moderators in the community would affect the transition probability of a user. We control for these two factors separately in W_{it}, and present the results in Table D2 and Table D3, respectively. Finally, we estimate the model on weekly data and include the results in Table D4.

Table D1. Results of HMM on Daily Data for Day 301-500 Subsample			
Variable Name	***State 1 (Low Motivation)***	***State 2 (Medium Motivation)***	***State 3 (High Motivation)***
$\boldsymbol{X}_{it}$	β – Posterior Mean (Standard Deviation)		
c^x	-2.791*** (0.071)	-0.103 (0.091)	4.735*** (0.330)
$Matched\_tags_{it}$	0.015*** (0.001)	0.029*** (0.001)	0.046*** (0.002)
$Group\_size_t$	0.001*** (0.000)	0.006*** (0.000)	0.013*** (0.002)
$Tenure_{it}$	-0.001*** (0.000)	-0.004*** (0.000)	-0.01*** (0.001)
$Total\_answers_{i,t-1}$	0.0002*** (0.000)	-0.0003*** (0.000)	-0.001*** (0.000)
σ^2	1.009*** (0.005)		
$\boldsymbol{W}_{i,t-1}$	ξ – Posterior Mean (Standard Deviation)		
c^w	-1.752*** (0.035)	-0.644*** (0.043)	0.941*** (0.114)
$Answers\_received_{i,t-1}$	0.268*** (0.019)	0.049* (0.026)	-0.117 (0.077)
$Upvotes\_answer_{i,t-1}$	0.262*** (0.020)	0.117*** (0.014)	0.021 (0.028)
$Accepted\_answers_{i,t-1}$	0.562*** (0.037)	0.300*** (0.024)	0.084** (0.035)
$Badges_{i,t-1}$	0.250*** (0.053)	0.275*** (0.040)	-0.095* (0.058)
Initial Probability	0.797*** (0.017)	0.186*** (0.016)	0.017*** (0.004)
* $p < 0.1$, ** $p < 0.05$; *** $p < 0.01$.			

Table D2. Results of HMM after Controlling Moderator in *W*			
Variable Name	**State 1 (Low Motivation)**	**State 2 (Medium Motivation)**	**State 3 (High Motivation)**
$\boldsymbol{X}_{it}$	β – Posterior Mean (Standard Deviation)		
c^x	-3.075*** (0.069)	0.325*** (0.087)	7.054*** (0.330)
$Matched\_tags_{it}$	0.015*** (0.001)	0.023*** (0.001)	0.030*** (0.001)
$Group\_size_t$	0.003*** (0.000)	0.002*** (0.000)	-0.010*** (0.002)
$Tenure_{it}$	-0.0004* (0.000)	-0.006*** (0.000)	-0.016*** (0.001)
$Total\_answers_{i,t-1}$	0.0005*** (0.000)	0.001*** (0.000)	0.003*** (0.000)
σ^2	1.010*** (0.006)		
$\boldsymbol{W}_{i,t-1}$	ξ – Posterior Mean (Standard Deviation)		
c^w	-1.655*** (0.018)	-0.581*** (0.026)	1.069*** (0.127)
$Answers\_received_{i,t-1}$	0.214*** (0.017)	0.021 (0.023)	0.013 (0.054)
$Upvotes\_answer_{i,t-1}$	0.238*** (0.023)	0.101*** (0.022)	0.015 (0.059)
$Accepted\_answers_{i,t-1}$	0.588*** (0.038)	0.221*** (0.036)	0.060 (0.096)
$Badges_{i,t-1}$	0.400*** (0.033)	0.198*** (0.033)	0.026 (0.063)
$Moderator_{i,t-1}$	-0.115 (0.086)	0.284*** (0.089)	0.803*** (0.145)
Initial Probability	0.758*** (0.014)	0.213*** (0.014)	0.029*** (0.005)
* $p < 0.1$, ** $p < 0.05$; *** $p < 0.01$.			

Table D3. Results of HMM after Controlling New Questions by Moderator in *W*

Variable Name	State 1 (Low Motivation)	State 2 (Medium Motivation)	State 3 (High Motivation)
$\boldsymbol{X}_{it}$	β – Posterior Mean (Standard Deviation)		
c^x	-2.990*** (0.080)	0.364*** (0.079)	7.368*** (0.365)
$Matched\_tags_{it}$	0.015*** (0.000)	0.023*** (0.001)	0.029*** (0.001)
$Group\_size_t$	0.002*** (0.001)	0.002*** (0.000)	-0.011*** (0.002)
$Tenure_{it}$	-0.0005** (0.000)	-0.006*** (0.000)	-0.015*** (0.001)
$Total\_answers_{i,t-1}$	0.0005*** (0.000)	0.001*** (0.000)	0.003*** (0.000)
σ^2	1.012*** (0.005)		
$\boldsymbol{W}_{i,t-1}$	ξ – Posterior Mean (Standard Deviation)		
c^w	-1.666*** (0.020)	-0.567*** (0.027)	1.442*** (0.246)
$Answers\_received_{i,t-1}$	0.220*** (0.016)	0.014 (0.023)	0.003 (0.059)
$Upvotes\_answer_{i,t-1}$	0.234*** (0.015)	0.115*** (0.014)	0.028 (0.027)
$Accepted\_answers_{i,t-1}$	0.592*** (0.035)	0.253*** (0.026)	0.075** (0.036)
$Badges_{i,t-1}$	0.408*** (0.034)	0.221*** (0.038)	-0.049 (0.064)
$New\_q\_moderators_{i,t-1}$	-0.002 (0.005)	-0.003 (0.009)	0.001 (0.038)
Initial Probability	0.758*** (0.016)	0.214*** (0.015)	0.028*** (0.005)
* $p < 0.1$, ** $p < 0.05$; *** $p < 0.01$.			

Table D4. Results of HMM on Weekly Data for Day 101-300 Subsample

Variable Name	State 1 (Low Motivation)	State 2 (Medium Motivation)	State 3 (High Motivation)
$\boldsymbol{X}_{it}$	β – Posterior Mean (Standard Deviation)		
c^x	-0.986*** (0.116)	10.566*** (0.772)	37.557*** (2.661)
$Matched\_tags_{it}$	0.003*** (0.000)	0.014*** (0.001)	0.031*** (0.001)
$Group\_size_t$	0.002*** (0.001)	-0.062*** (0.005)	-0.213*** (0.016)
$Tenure_{it}$	-0.002*** (0.000)	-0.029*** (0.001)	-0.070*** (0.003)
$Total\_answers_{i,t-1}$	0.002*** (0.000)	0.013*** (0.000)	0.024*** (0.001)
σ^2	2.028*** (0.030)		
$\boldsymbol{W}_{i,t-1}$	ξ – Posterior Mean (Standard Deviation)		
c^w	-1.835*** (0.043)	-0.793*** (0.066)	-0.124*** (0.303)
$Answers\_received_{i,t-1}$	0.072*** (0.012)	0.023 (0.015)	0.021 (0.035)
$Upvotes\_answer_{i,t-1}$	0.069*** (0.017)	0.018 (0.017)	0.013 (0.027)
$Accepted\_answers_{i,t-1}$	0.110*** (0.038)	0.056*** (0.018)	0.128 (0.083)
$Badges_{i,t-1}$	0.183*** (0.033)	0.144*** (0.036)	-0.054 (0.044)
Initial Probability	0.769*** (0.045)	0.192*** (0.041)	0.039*** (0.011)
* $p < 0.1$, ** $p < 0.05$; *** $p < 0.01$.			

References

Albert, J. H., and Chib, S. 1993. "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Association* (88:422), pp. 669-679.

Kim, C.-J., and Nelson, C. R. 1999. *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*, Cambridge, MA: The MIT Press.

MacDonald, I. L., and Zucchini, W. 1997. *Hidden Markov and Other Models for Discrete-Valued Time Series*, Boca Raton, FL: CRC Press.

Netzer, O., Lattin, J. M., and Srinivasan, V. 2008. "A Hidden Markov Model of Customer Relationship Dynamics," *Marketing Science* (27:2), pp. 185-204.

Singh, P. V., Tan, Y., and Youn, N. 2011. "A Hidden Markov Model of Developer Learning Dynamics in Open Source Software Projects," *Information Systems Research* (22:4), pp. 790-807.

Smith, A., Naik, P. A., and Tsai, C.-L. 2006. "Markov-Switching Model Selection Using Kullback-Leibler Divergence," *Journal of Econometrics* (134:2), pp. 553-577.

Yan, L., and Tan, Y. 2014. "Feeling Blue? Go Online: An Empirical Study of Social Support Among Patients," *Information Systems Research* (25:4), pp. 690-709.